

# Esame 11/09/2019

Andrea Passerini  
andrea.passerini@unitn.it

Informatica

## Programma python

Scrivere un programma che prenda in ingresso un file con i dati di espressione genica binarizzata (sovraespresso o meno) di pazienti affetti da leucemia di tipo AML o ALL e:

- Calcoli il coefficiente di correlazione tra i valori di espressione di ciascun gene ed una patologia di tipo AML
- Stampi la lista dei geni ordinati in base al loro coefficiente di correlazione con la patologia AML

# Esempio file di input

```
> cat gene_exp.txt
ATP2B4,NAP1L1,MDK,PCCB,MDS1,AML
no,no,yes,no,no,no
no,no,no,yes,yes,no
no,no,no,yes,yes,no
no,no,yes,no,yes,no
no,no,no,no,yes,yes
no,yes,no,yes,yes,no
no,no,no,yes,yes,no
no,no,no,yes,yes,yes
yes,no,yes,yes,yes,no
no,no,no,yes,yes,yes
no,yes,no,yes,no,no
yes,no,no,yes,yes,yes
yes,no,yes,yes,yes,no
no,no,no,yes,yes,no
```

# Esempio esecuzione

```
> python correlation2AML.py  
Inserire nome file: gene_exp.txt  
gene      corrcoef  
MDS1      0.26  
ATP2B4    0.06  
PCCB      -0.06  
NAP1L1    -0.26  
MDK       -0.40
```

## Programma python: suggerimento

Si possono implementare 5 funzioni separate:

- 1 una che legga il file dati e restituisca l'intestazione e una matrice di dati
- 2 una che data una matrice di dati e un indice di colonna, estragga la colonna corrispondente all'indice, convertendo i valori in numerici (0 per "no", 1 per "yes")
- 3 una che dati header e matrice di dati, per ogni gene estragga la colonna corrispondente e calcoli la sua correlazione con la patologia AML (ultima colonna della matrice). La funzione restituirà una lista di coppie gene-correlazione.
- 4 una che data la lista di correlazioni la stampi in ordine decrescente
- 5 una (o un main) che realizzi il programma richiesto usando le funzioni di cui sopra

## Programma python: NOTA

Per calcolare il coefficiente di correlazione tra due liste, si può usare la funzione `corrcoef` del modulo `numpy`

```
from numpy import corrcoef

a = [1,0,1,0,0,1]
b = [0,1,1,0,0,1]

corrcoef(a,b)[0,1]
# corrcoef restituisce una matrice di correlazione
# con [0,1] si prende la correlazione tra a e b
```

## Esercizi da linea di comando

Identificare (con un singolo comando) le sequenze che cominciano per metionina (M), finiscono per acido aspartico o glutammico (D o E), e sono lunghe almeno 200 residui

## Risultato atteso

1g6wD.f

1gp5A.f

### Esercizi da linea di comando

Calcolare quanti pazienti affetti da AML hanno un profilo identico tra loro.

### Risultato atteso

2



## Modalita' di esecuzione e consegna

- 1 Avviare la macchina in modalita' `ESAME`
- 2 Autenticarsi con nome utente `sci-esame` e password fornita dal docente
- 3 Il testo del compito ed i file necessari si trovano in una cartella `Testo` sul Desktop
- 4 Realizzare il programma python come file `programma.py` e scrivere gli esercizi da linea di comando in un file di testo `linea_di_comando.txt`
- 5 Creare sul Desktop una cartella con *nome\_cognome* e metterci i due file realizzati.
- 6 Eseguire il logout ma NON spegnere la macchina